# Accounting for model biases in 4D-Var

**Yannick TRÉMOLET - ECMWF**

**Workshop on Reanalysis - June 2006**

- Introduction
- Weak constraint 4D-Var
- Results
- Future work

# 4D Variational Data Assimilation

Variational data assimilation is based on the minimisation of:

$$
\begin{aligned}
J(x) &= \frac{1}{2}[\mathcal{H}(x) - y]^T R^{-1} [\mathcal{H}(x) - y] \\
&+ \frac{1}{2}(x_0 - x_b)^T B^{-1} (x_0 - x_b) + \frac{1}{2}\mathcal{F}(x)^T C^{-1} \mathcal{F}(x)
\end{aligned}
$$

- $x$ is the 4D state of the atmosphere over the assimilation window.
- $\mathcal{H}$ is a 4D observation operator, accounting for the time dimension.
- $\mathcal{F}$ represents the remaining theoretical knowledge after background information has been accounted for (balance, DFI...).
- Control variable reduces to $x_0$ using the relation: $x_i = \mathcal{M}_i(x_{i-1})$.
- Used in operational 4D-Var implementations.
- Model $\mathcal{M}$ verified exactly although it is not perfect...

# Weak constraint 4D-Var

- The model can be imposed as a constraint in the cost function, in the same way as other sources of information:

$$\mathcal{F}_i(x) = x_i - \mathcal{M}_i(x_{i-1})$$

- Model error $\eta$ is defined as: $\eta_i = x_i - \mathcal{M}_i(x_{i-1})$
- The cost function becomes:

$$\begin{aligned} J(x) &= \frac{1}{2}\sum_{i=0}^{n}[\mathcal{H}(x_i) - y_i]^T R_i^{-1}[\mathcal{H}(x_i) - y_i] \\ &+ \frac{1}{2}(x_0 - x_b)^T B^{-1}(x_0 - x_b) + \frac{1}{2}\sum_{i=1}^{n}\eta_i^T Q_i^{-1}\eta_i \end{aligned}$$

- Model error covariance matrix $Q$ has to be defined.
- Strong constraint 4D-Var is $\mathcal{F}_i(x) \equiv 0$ i.e. $\eta \equiv 0$ (perfect model).

# Control Variable in 4D-Var

$$J(x) = \frac{1}{2}(x_0 - x_b)^T B^{-1}(x_0 - x_b) + \frac{1}{2}\sum_{i=0}^{n}[\mathcal{H}(x_i) - y_i]^T R_i^{-1}[\mathcal{H}(x_i) - y_i]$$

$$+ \quad \frac{1}{2}\sum_{i=1}^{n}[x_i - \mathcal{M}_i(x_{i-1})]^T Q_i^{-1}[x_i - \mathcal{M}_i(x_{i-1})]$$

| 4D-Var | 4D-Var$_x$ | 4D-Var$_\eta$ | 4D-Var$_\beta$ |
|---|---|---|---|
| $x_0$ | $x$ | $x_0, \eta$ | $x_0, \beta$ |
| $x_i = \mathcal{M}_i(x_{i-1})$ | $x_i \approx \mathcal{M}_i(x_{i-1})$ | $x_i = \mathcal{M}_i(x_{i-1}) + \eta_i$ | $x_i = \mathcal{M}_{i,0}(x_0) + \beta_i$ |
| $\Downarrow$ | $\Downarrow$ | $\Downarrow$ | $\Downarrow$ |
| 3D Initial Condition | 4D Model Trajectory | 3D I.C. + Model Error Forcing | 3D I.C. + Model Bias |

## Model Error Forcing Control Variable

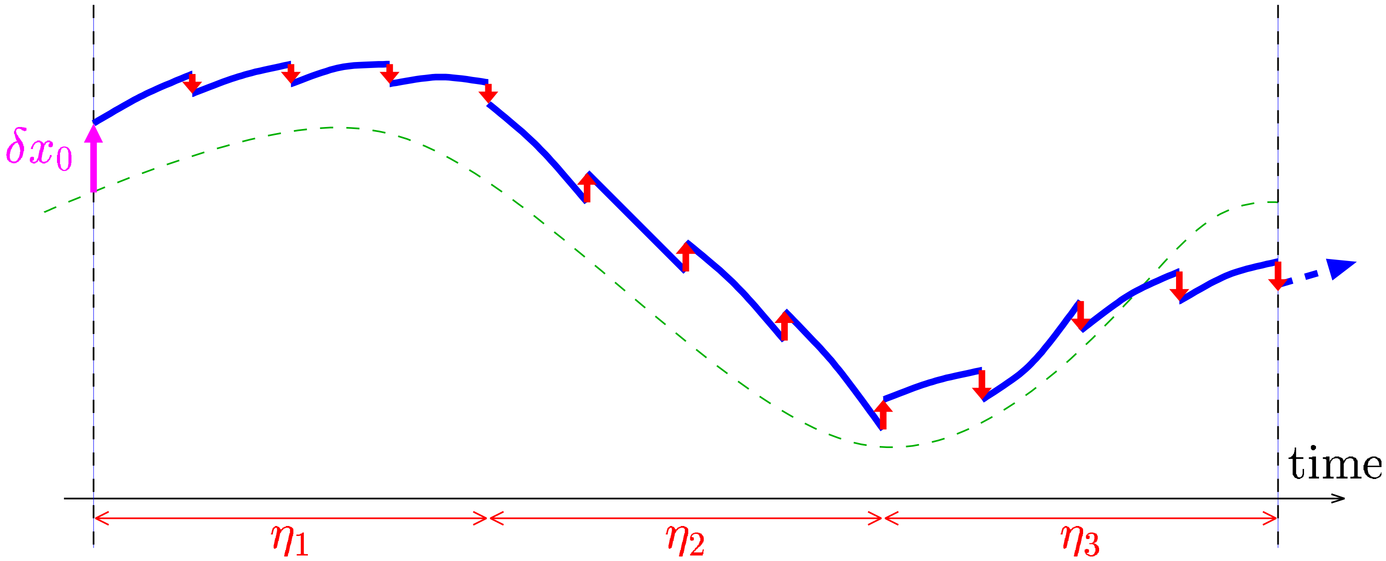
# 4D-Var with Model Error Forcing

$$J(x_0, \eta) = \frac{1}{2} \sum_{i=0}^{n} [\mathcal{H}(x_i) - y_i]^T R_i^{-1} [\mathcal{H}(x_i) - y_i] + \frac{1}{2}(x_0 - x_b)^T B^{-1} (x_0 - x_b) + \eta^T Q^{-1} \eta$$

$$\text{with } x_i = \mathcal{M}_i(x_{i-1}) + \eta_i .$$

- The *usual* model error term in 4D-Var.
- $\eta_i$ is a 3D atmospheric state,
- $\eta_i$ represents the instantaneous model error,
- $\eta$ is constrained by the fact that it is propagated by the model.

# 4D-Var with Model Error Forcing



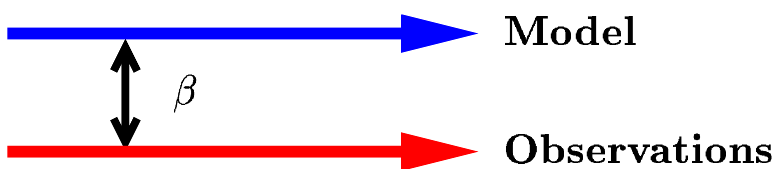

- TL and AD models can be used with little modification,
- Information is propagated between obervations and IC control variable by TL and AD models.

## Model Bias Control Variable

# 4D-Var with Model Bias

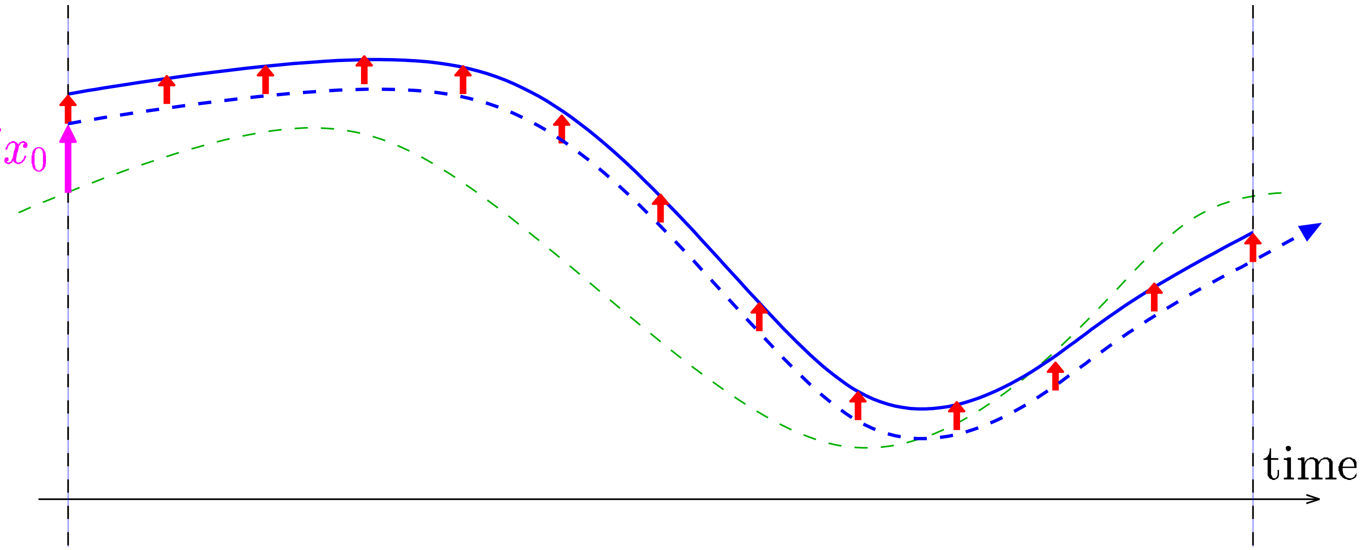
$$
\begin{aligned}
J(x_0, \beta) \quad &= \quad \frac{1}{2} \sum_{i=0}^{n} [\mathcal{H}(x_i^m + \beta_i) - y_i]^T R_i^{-1} [\mathcal{H}(x_i^m + \beta_i) - y_i] \\
&+ \quad \frac{1}{2} (x_0 - x_b)^T B^{-1} (x_0 - x_b) + \beta^T Q_\beta^{-1} \beta
\end{aligned}
$$

$$
\text{with } x_i^m = \mathcal{M}_{i,0}(x_0).
$$




- $\beta_i$ is a 3D atmospheric state,
- The model is not perturbed,
- $\beta$ sees global (model – all observations) bias,
- Does not correct for bias of one subset of observations against another subset of observations.

# 4D-Var with Model Bias




- Bias added to forecast at post-processing stage,
- Makes sense if $\beta$ is slowly varying or constant ($\beta_i = \beta$),
- Information is propagated between obervations and IC control variable by TL and AD models (not modified).
- Model bias is represented by additional parameters not entering model equations,
- Optimisation problem is very similar to strong constraint 4D-Var.

## Model State Control Variable

# Model State Control Variable

- Use $\{x_i\}_{i=0,\ldots,n}$ as the control variable.
- Incremental cost function:

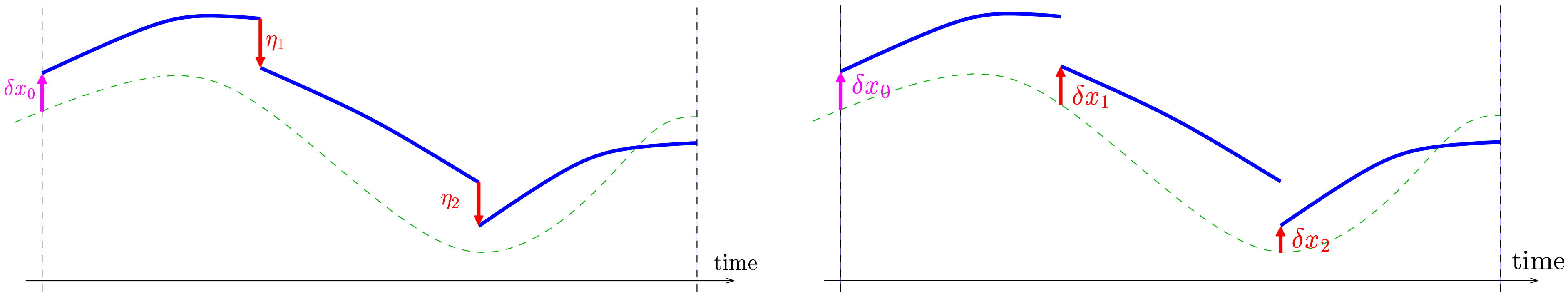
$$
\begin{aligned}
J(\delta x) &= \frac{1}{2}(\delta x_0 - b)^T B^{-1}(\delta x_0 - b) + \frac{1}{2}\sum_{i=0}^{n}(H\delta x_i - d_i)^T R_i^{-1}(H\delta x_i - d_i) \\
&+ \frac{1}{2}\sum_{i=1}^{n}(q_i + M_{i-1}\delta x_{i-1} - \delta x_i)^T Q_i^{-1}(q_i + M_{i-1}\delta x_{i-1} - \delta x_i)
\end{aligned}
$$

where $b = x^g - x_b$, $d_i = \mathcal{H}(x_i^g) - y_i$ and $q_i = \mathcal{M}_{i-1}(x_{i-1}^g) - x_i^g$.



ECMWF

# Model State Control Variable




Forcing Control Variable    Model State Control Variable

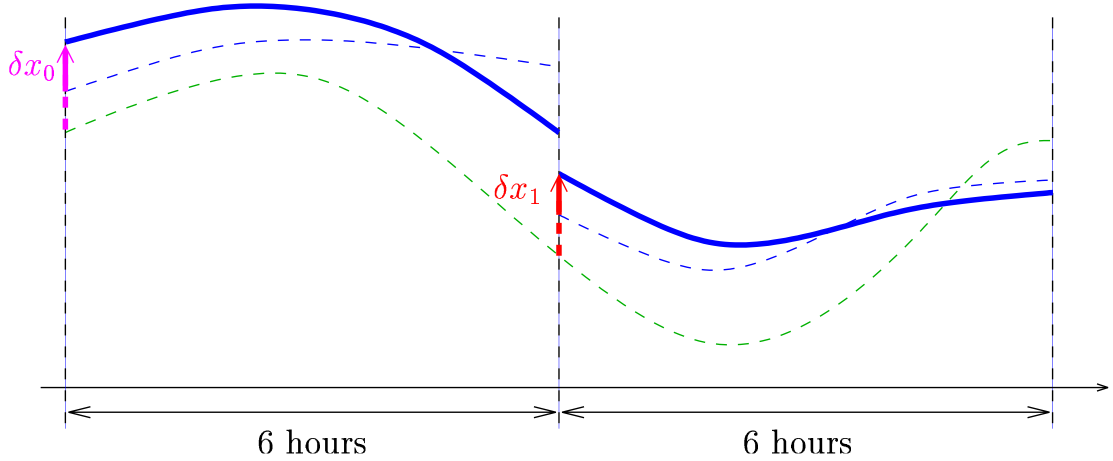
- Model integrations within each time-step (or sub-window) are independent:
  - Information is not propagated across sub-windows by TL/AD models,
  - Natural parallel implementation (in theory...).
- Tangent linear and adjoint models:
  - can be used without modification,
  - propagate information between observations and control variable within each sub-window.



ECMWF

# Weak Constraint 4D-Var: Examples

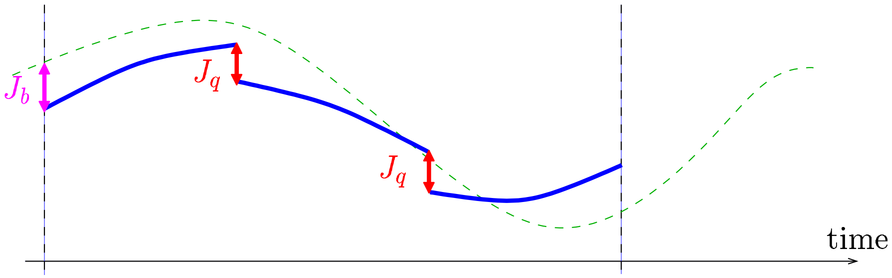
- 6-hour sub-windows:



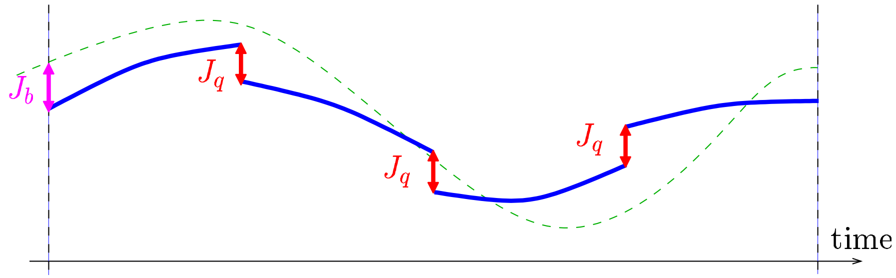

  - Better than 6-hour 4D-Var: two cycles are coupled through $J_q$,
  - Better than 12-hour 4D-Var: more information (imperfect model), more control,

- One time step sub-windows:
  - Each assimilation problem is instantaneous = 3D-Var,
  - Equivalent to a string of 3D-Var problems coupled together and solved as a single minimisation problem,
  - Approximation can be extended to non instantaneous sub-windows.

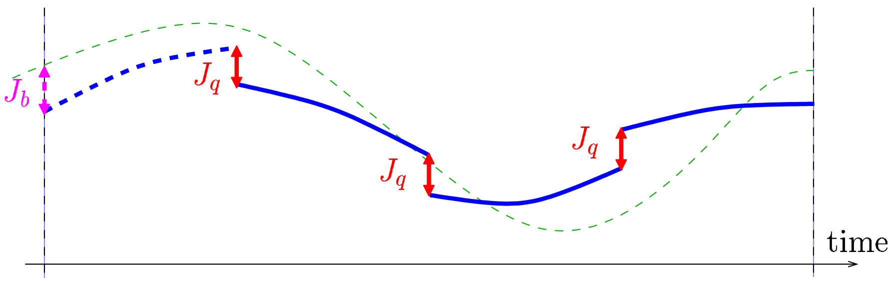
# Weak Constraint 4D-Var: Sliding Window



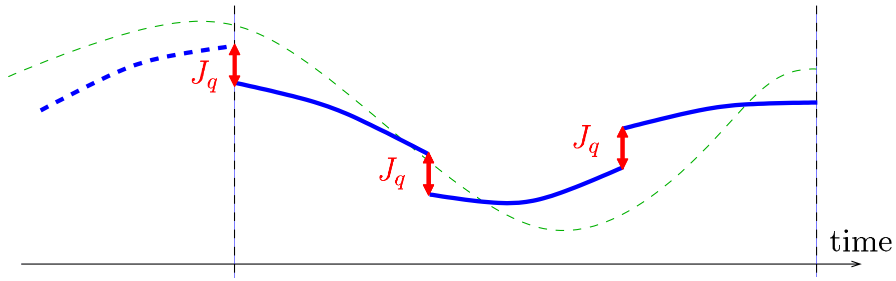

(1) Weak constraint 4D-Var




(2) Extended window




(3) Initial term has converged



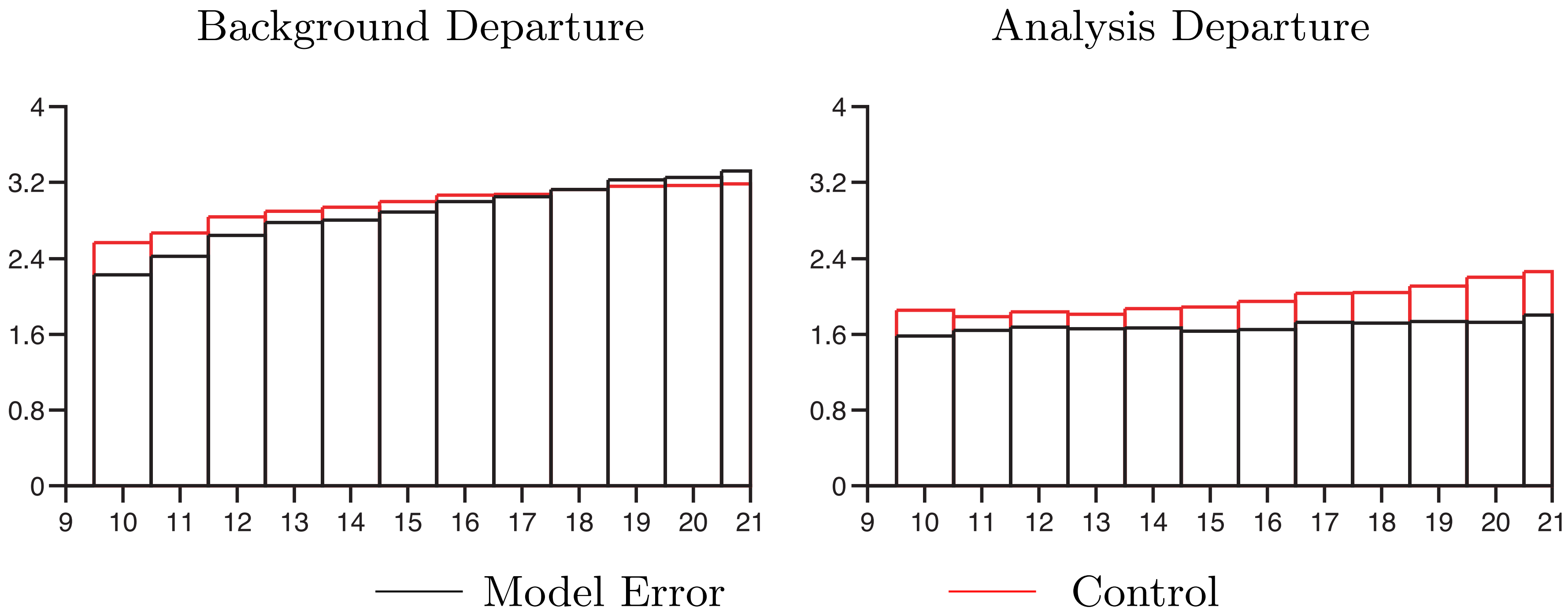

(4) Assimilation window is moved forward

- This implementation is an approximation of weak contraint 4D-Var with an assimilation window that extends indefinitely in the past...
- ...which is equivalent to a Kalman smoother that has been running indefinitely (M. Fisher).

# Results: Constant Model Error Forcing

# Results: Fit to observations

```
AMprofiler-windspeed Std Dev N.Amer
```



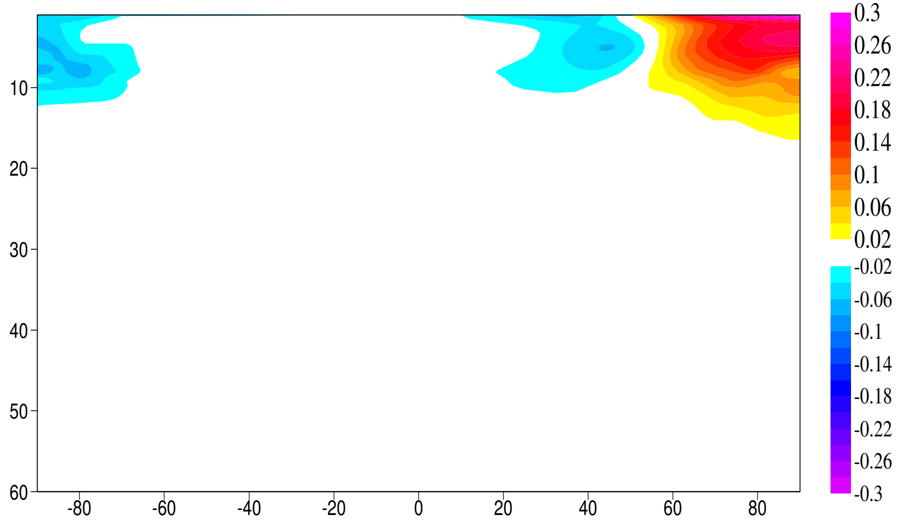

- Fit to observations is more uniform over the assimilation window.
- Background fit improved only at the start: error varies in time ?

# Model Error Forcing

Zonal Mean Temperature

July 2004

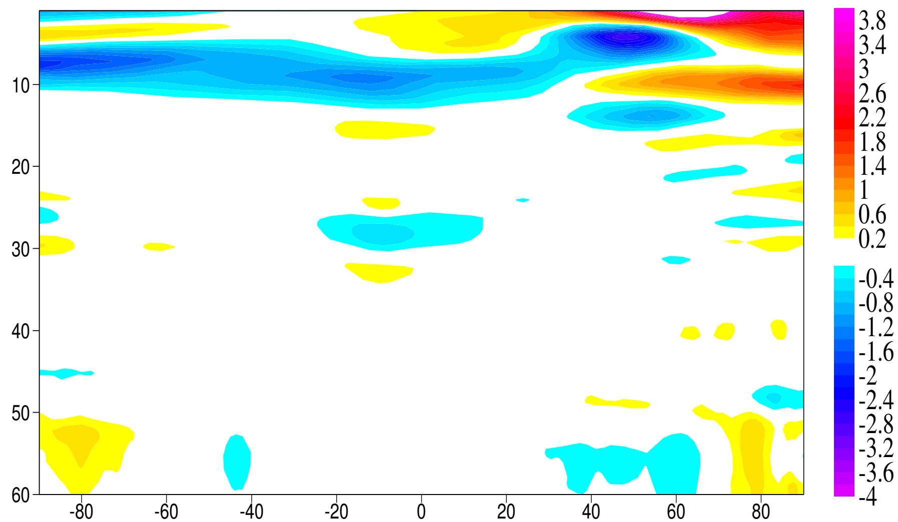
M.E. Forcing ⟶

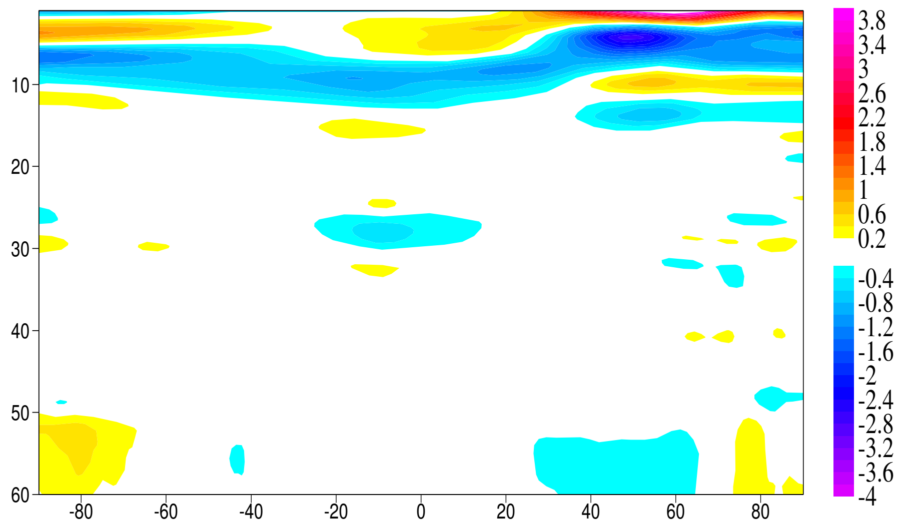



M.E. Mean Increment ↘

Control Mean Increment

↓



# Mean Model Error Forcing

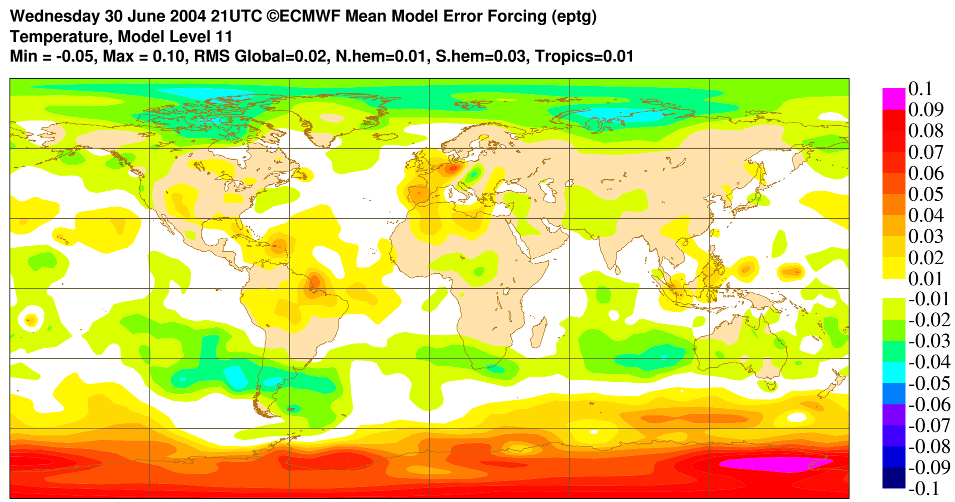
Temperature

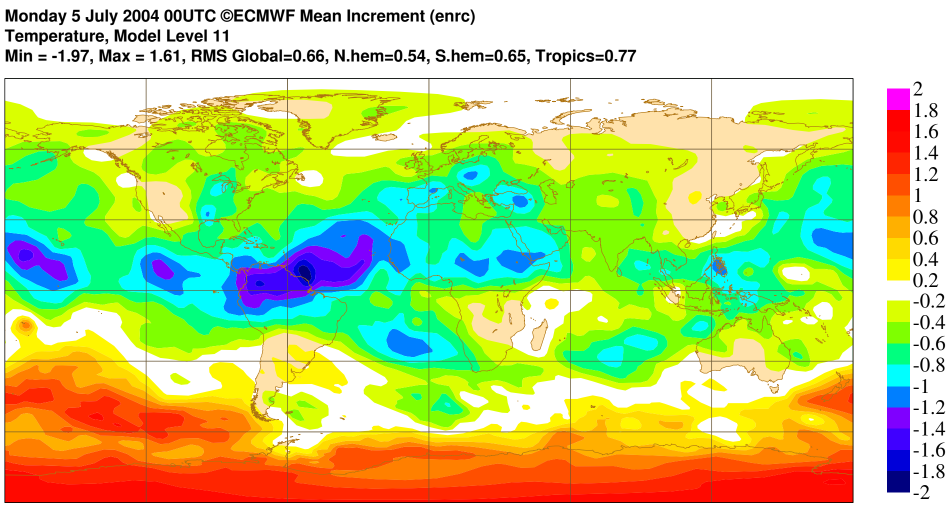
Model level 11 (≈5hPa)

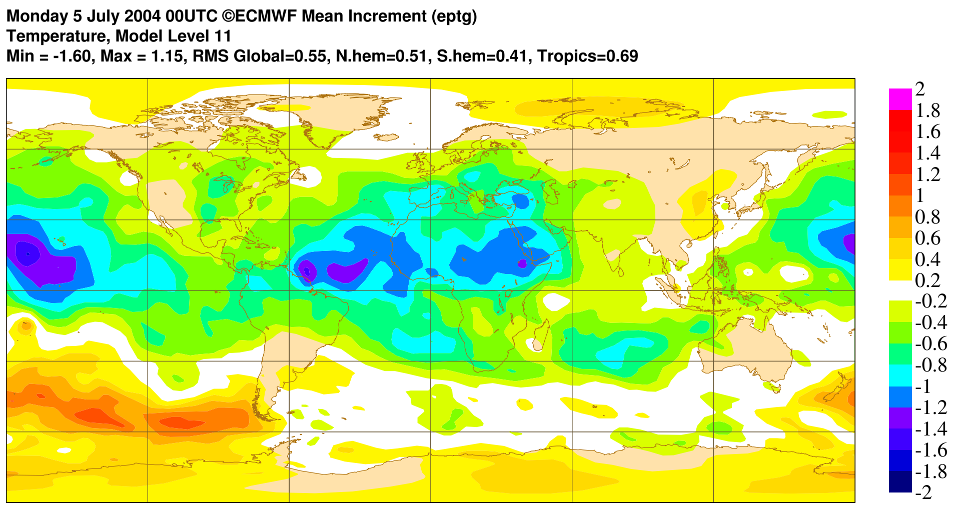
Mean M.E. Forcing ⟶

M.E. Mean Increment ↘

Control Mean Increment ↓

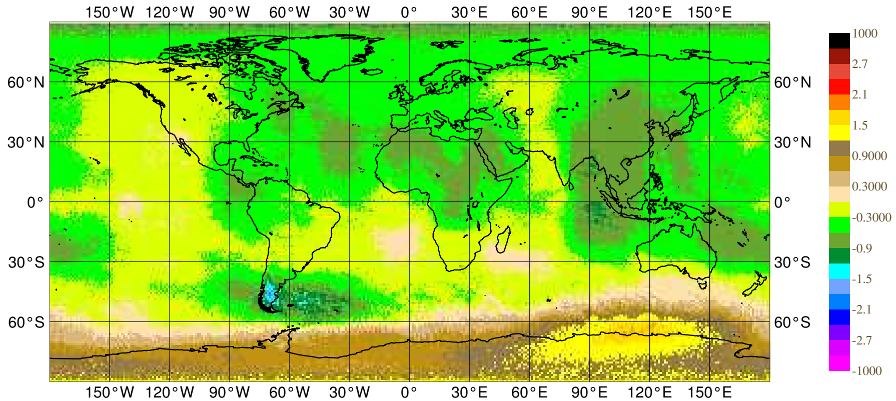


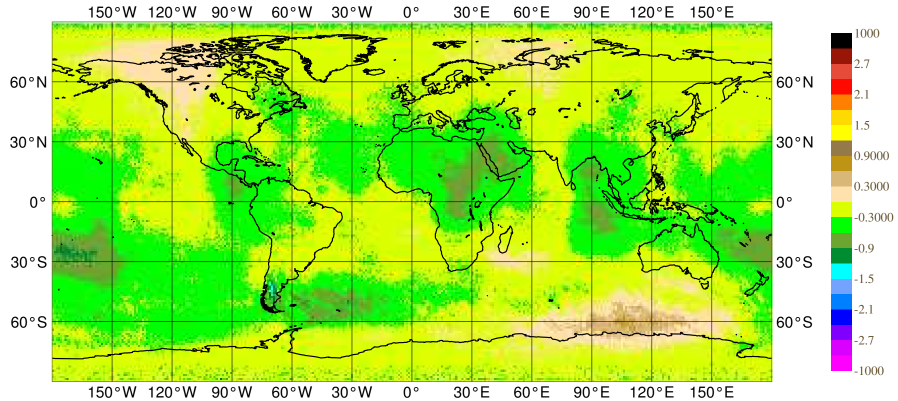



# AMSU-A FG Departures

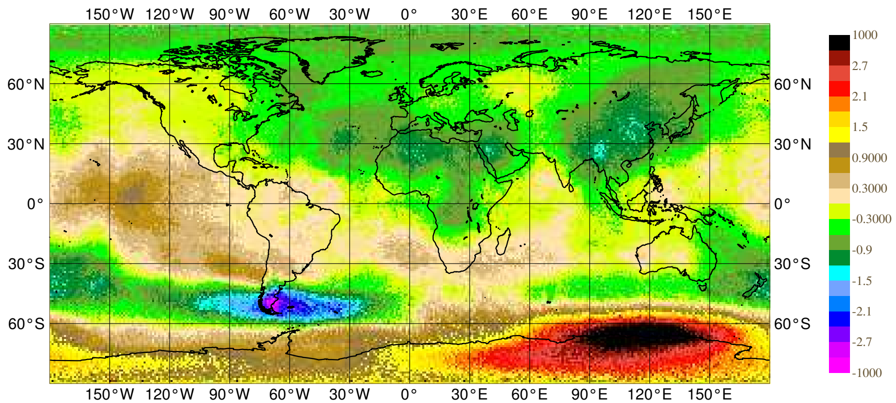
STATISTICS FOR RADIANCES FROM NOAA-16 / AMSU-A - 13
MEAN FIRST GUESS DEPARTURE (OBS-FG) (BCORR.) (CLEAR)
DATA PERIOD = 2004070200 - 2004073118 , HOUR = ALL
EXP = ENRC
Min: -1.9618 Max: 2.7 Mean: -0.169506




STATISTICS FOR RADIANCES FROM NOAA-16 / AMSU-A - 13
MEAN FIRST GUESS DEPARTURE (OBS-FG) (BCORR.) (CLEAR)
DATA PERIOD = 2004070100 - 2004073118 , HOUR = ALL
EXP = EPTG
Min: -1.6688 Max: 0.8 Mean: -0.231773

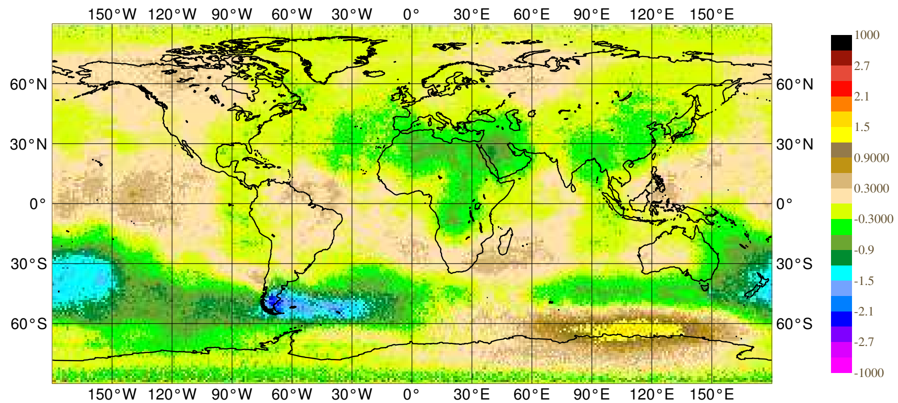



STATISTICS FOR RADIANCES FROM NOAA-16 / AMSU-A - 14
MEAN FIRST GUESS DEPARTURE (OBS-FG) (BCORR.) (CLEAR)
DATA PERIOD = 2004070200 - 2004073118 , HOUR = ALL
EXP = ENRC
Min: -3.3564 Max: 5.46 Mean: 0.006309




Control

STATISTICS FOR RADIANCES FROM NOAA-16 / AMSU-A - 14
MEAN FIRST GUESS DEPARTURE (OBS-FG) (BCORR.) (CLEAR)
DATA PERIOD = 2004070100 - 2004073118 , HOUR = ALL
EXP = EPTG
Min: -2.6 Max: 2.16 Mean: -0.111883

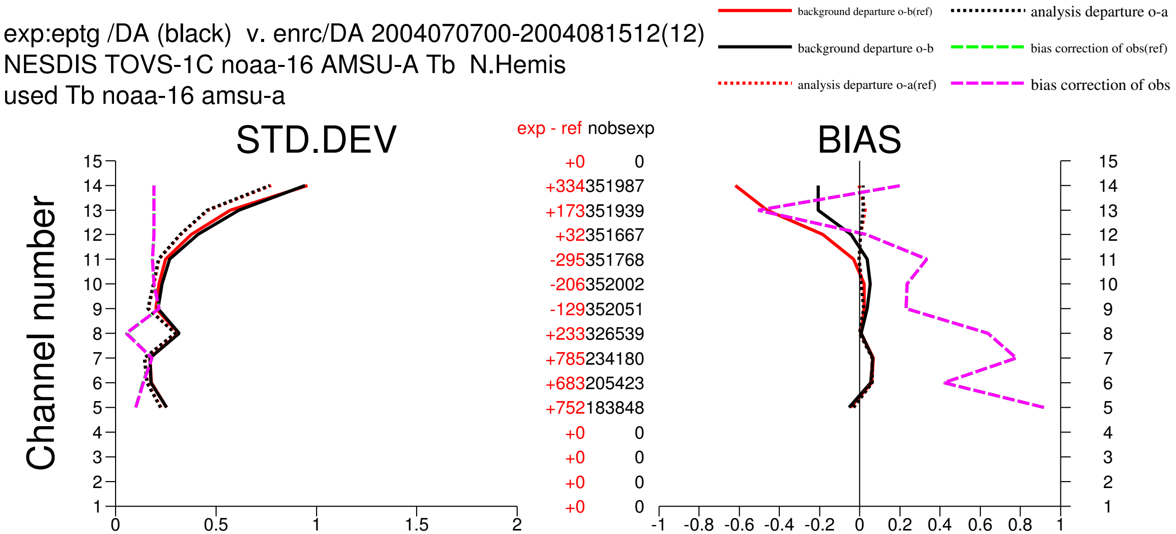


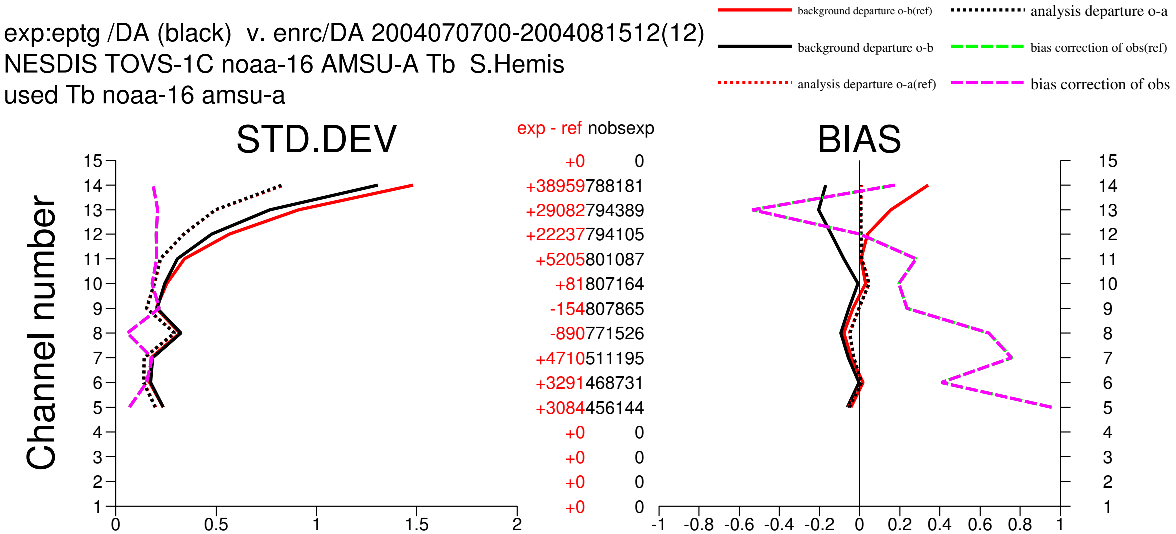

Model Error



ECMWF

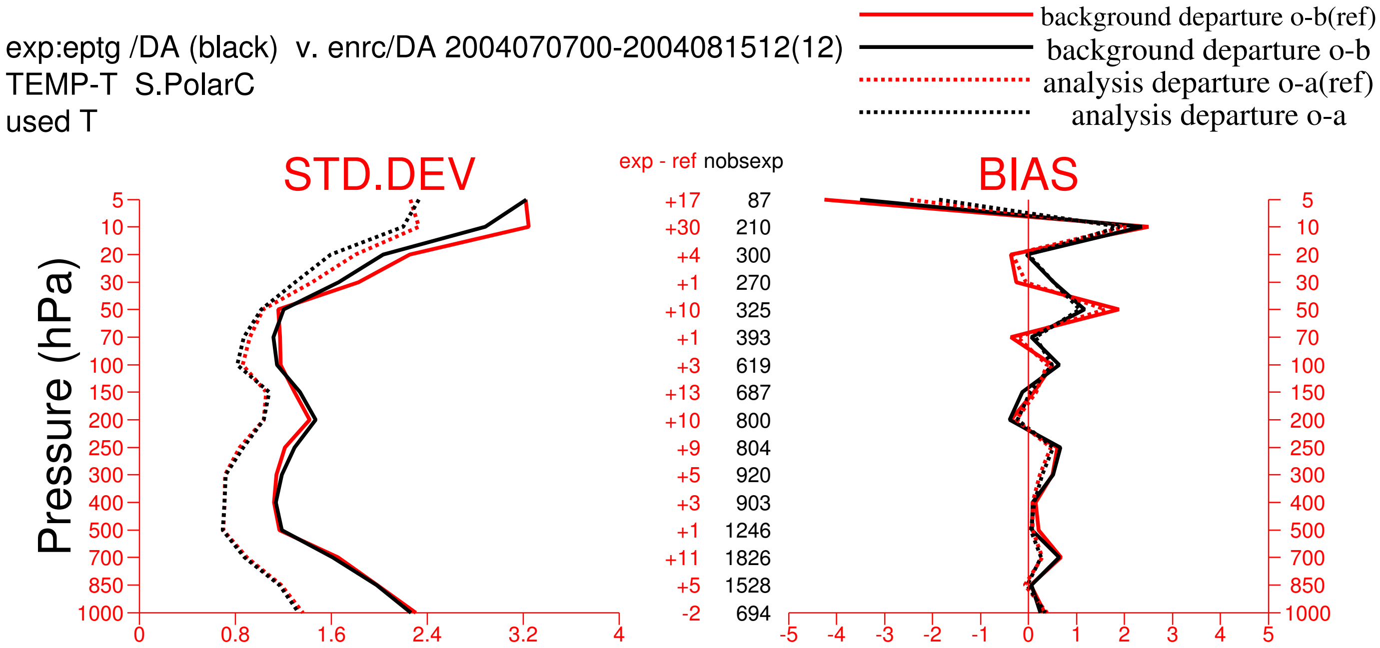
# AMSU-A Statistics




- More data is used,
- Bias is more uniform,
- BG std. dev. is reduced in SH.

# Fit to radiosonde data




- Oscillations in polar stratosphere are reduced,
- Std. deviation is reduced above 50 hPa (bg and an).

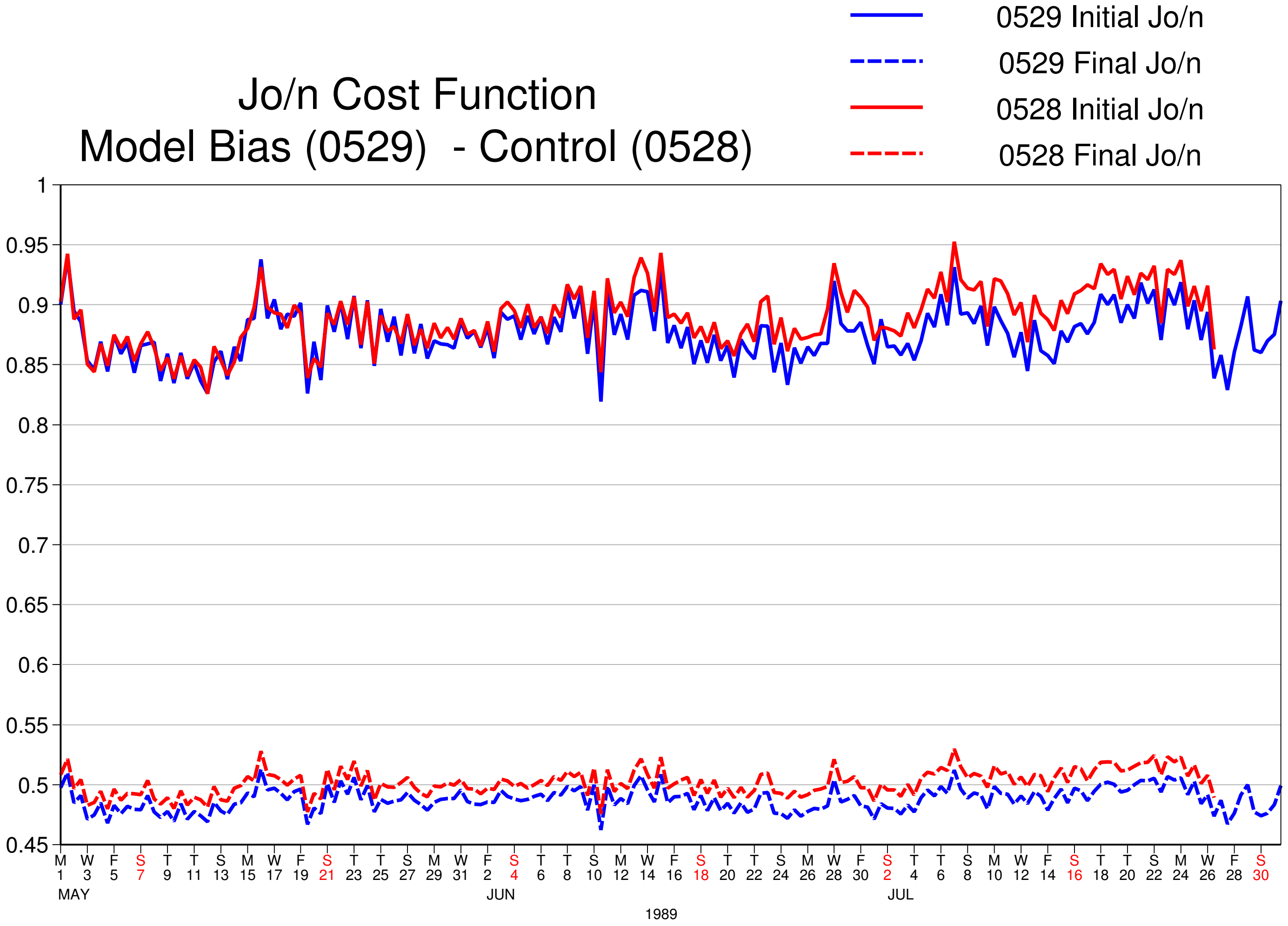
**One source of model error was corrected by the forcing term.**



ECMWF

# Results: Model Bias

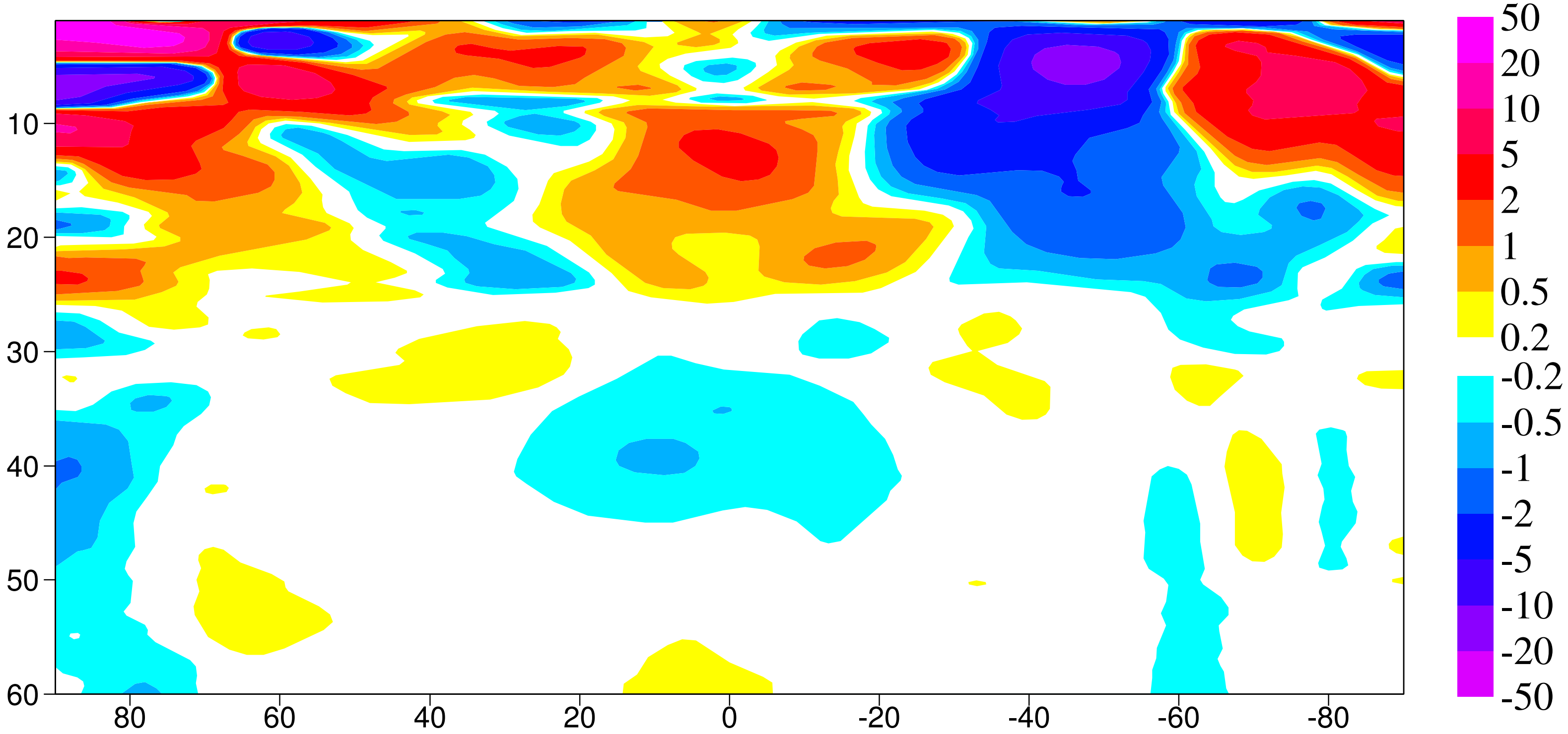
# Model Bias: Fit to Observations




Fit to analysis and background is improved.

# Model Bias

Average Model Bias - Temperature (K) - July 1989






ECMWF

# Mean Model Bias

Temperature

Model level 4 (≈1hPa)

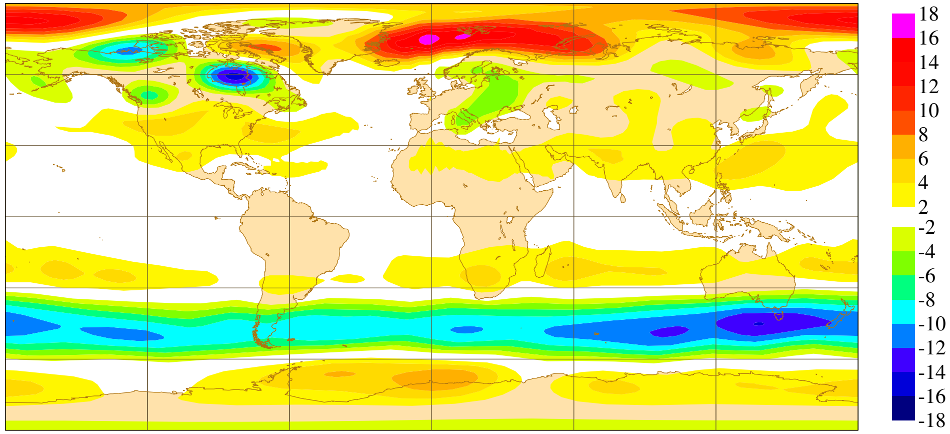
June 1989

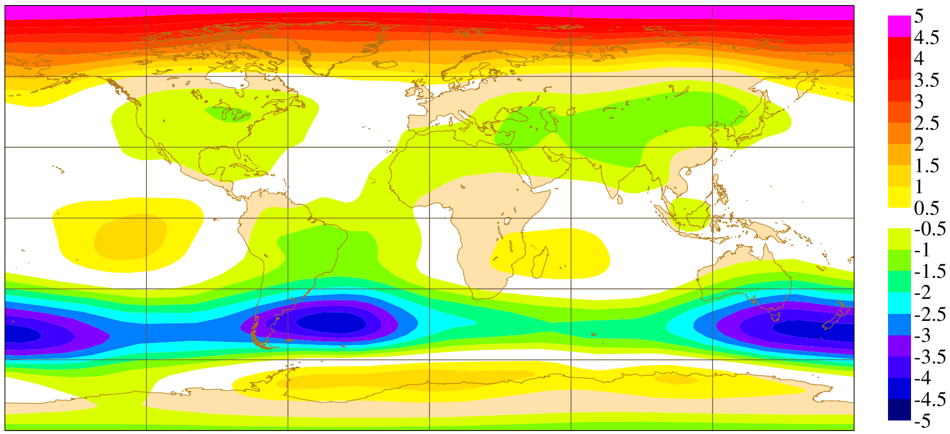
Mean Bias ⟶

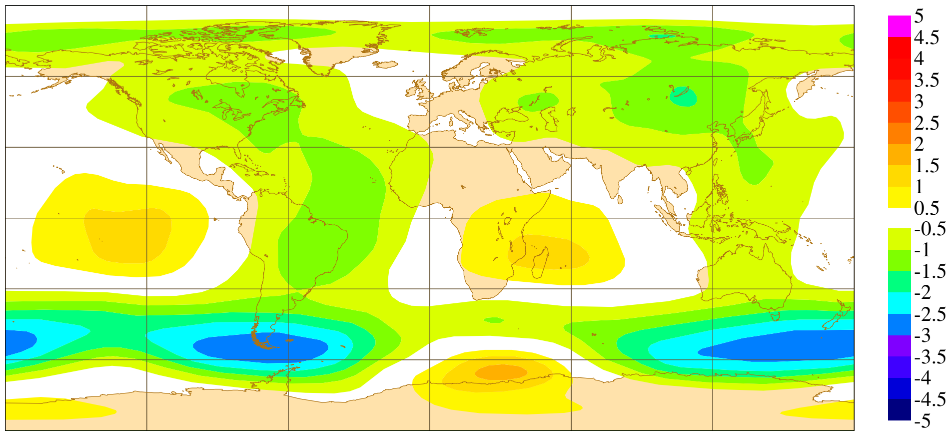
Mean Increment with Bias ↘

Control Mean Increment

↓



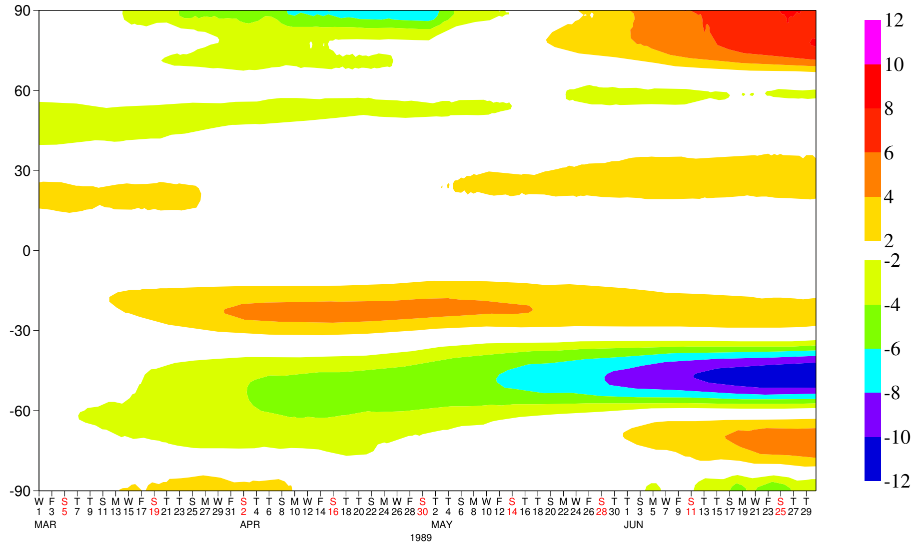



# Mean Model Bias

Temperature

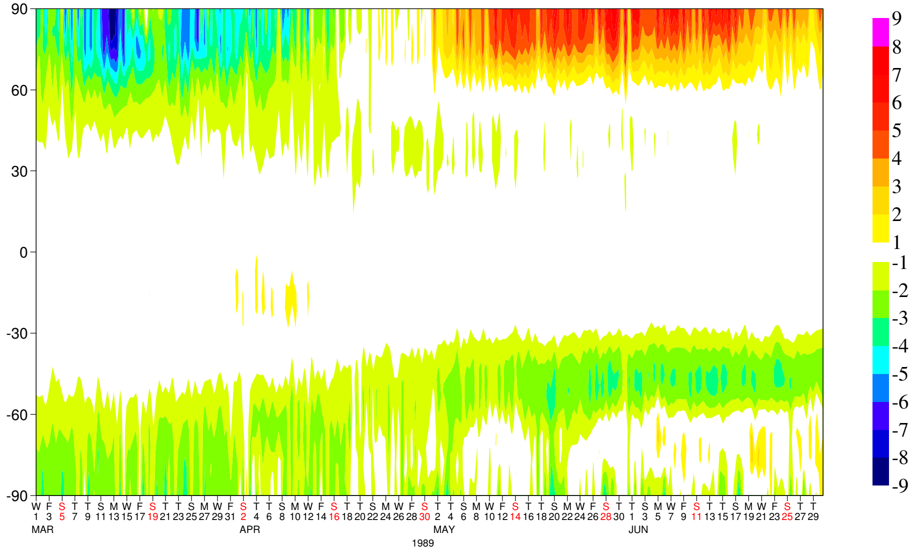
Model level 4 (≈1hPa)

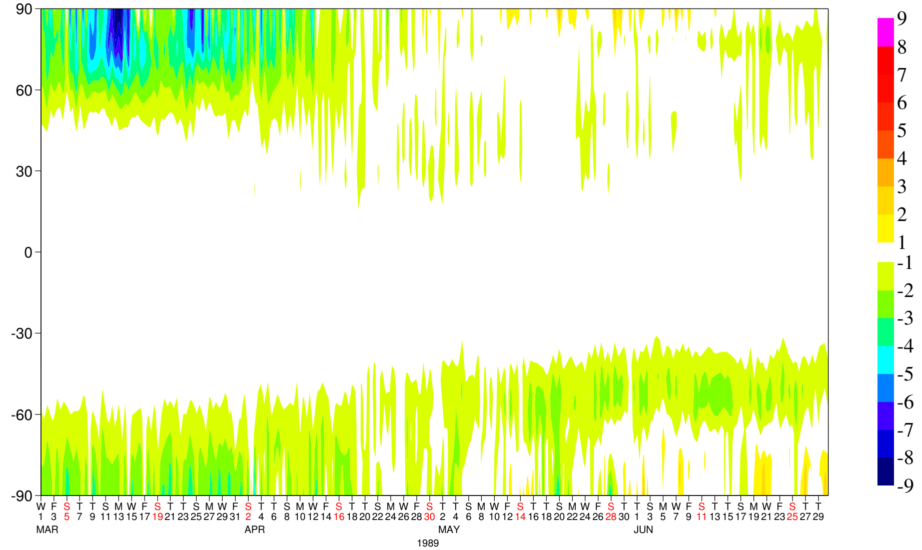
Bias



Increment with Bias

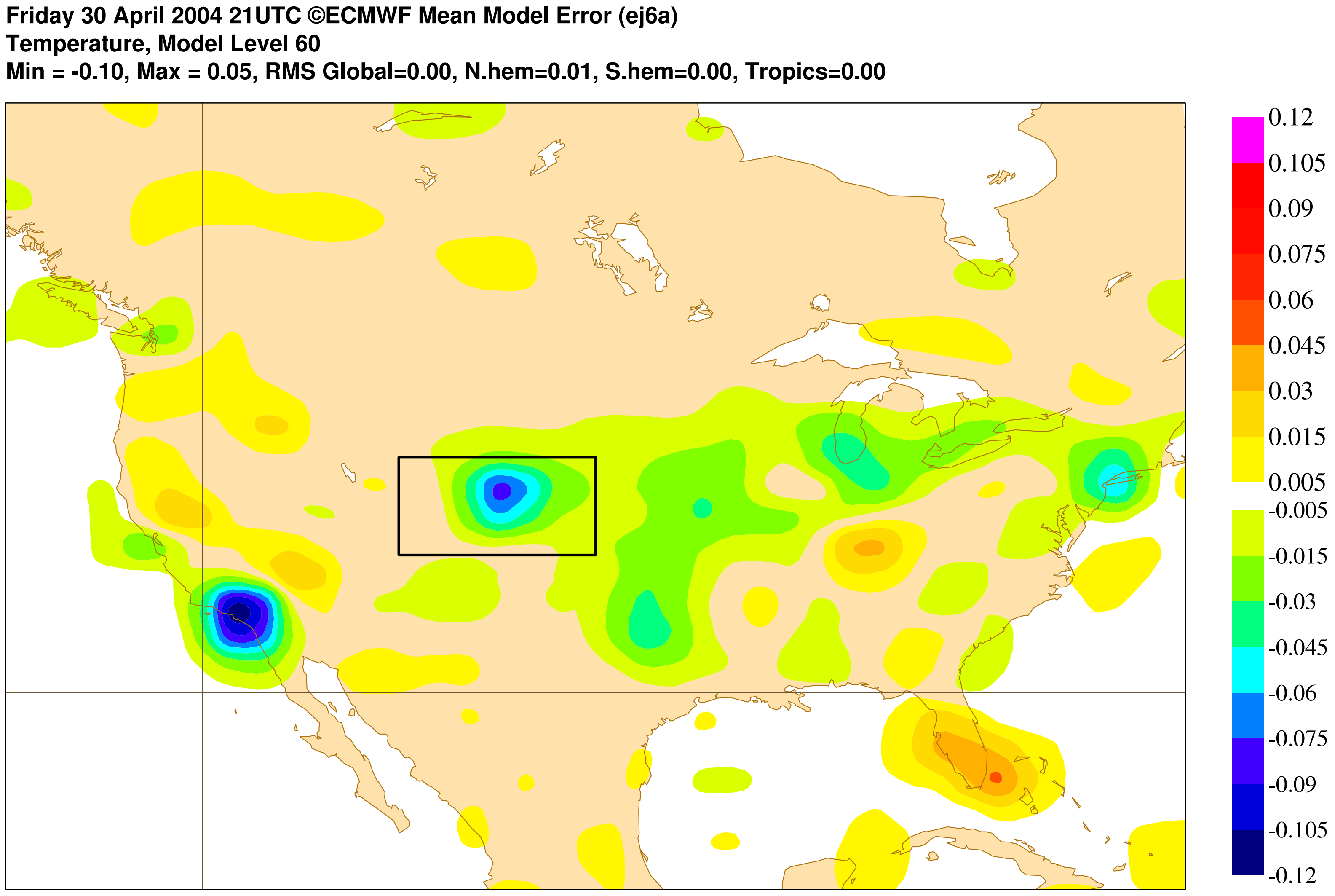
Control Increment



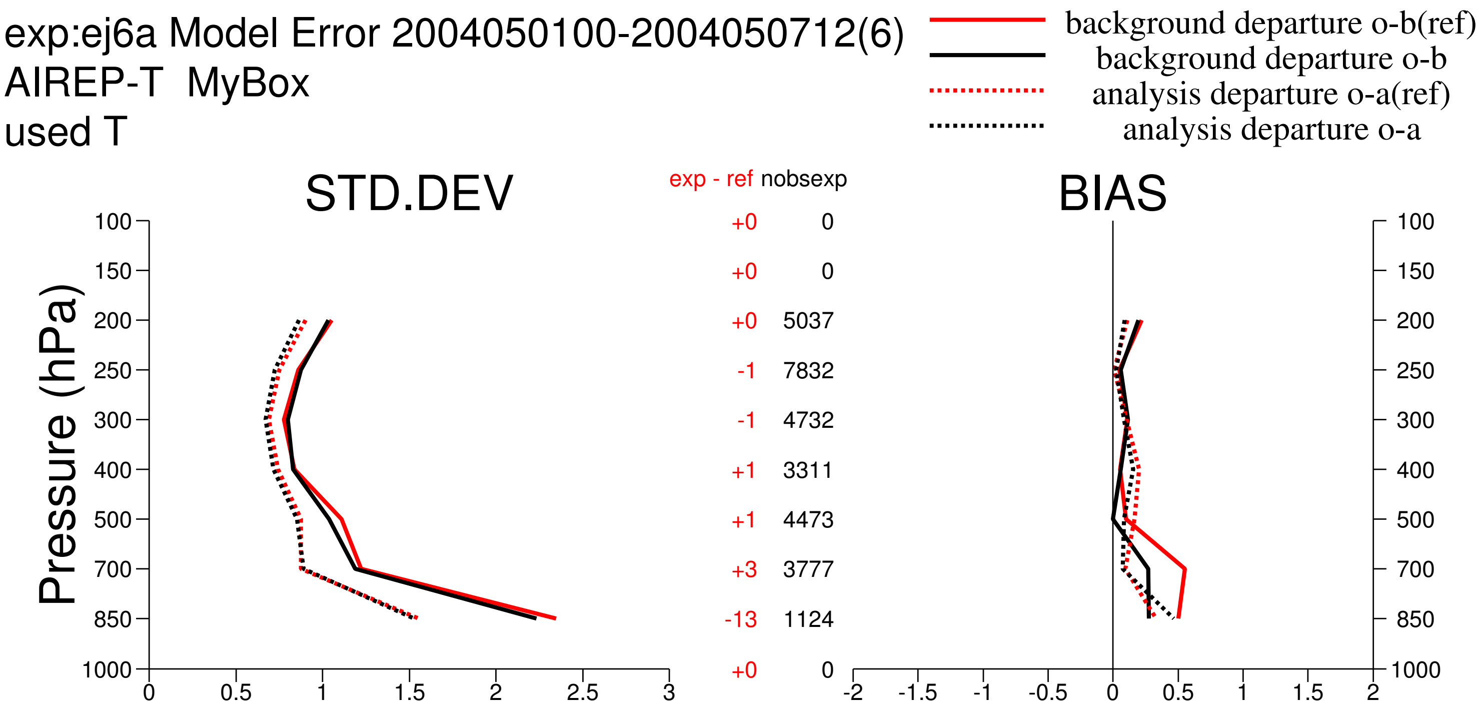


# Model Bias and Observation Bias

# Low Level Mean Model Error Forcing



# Fit to Observation with Model Error

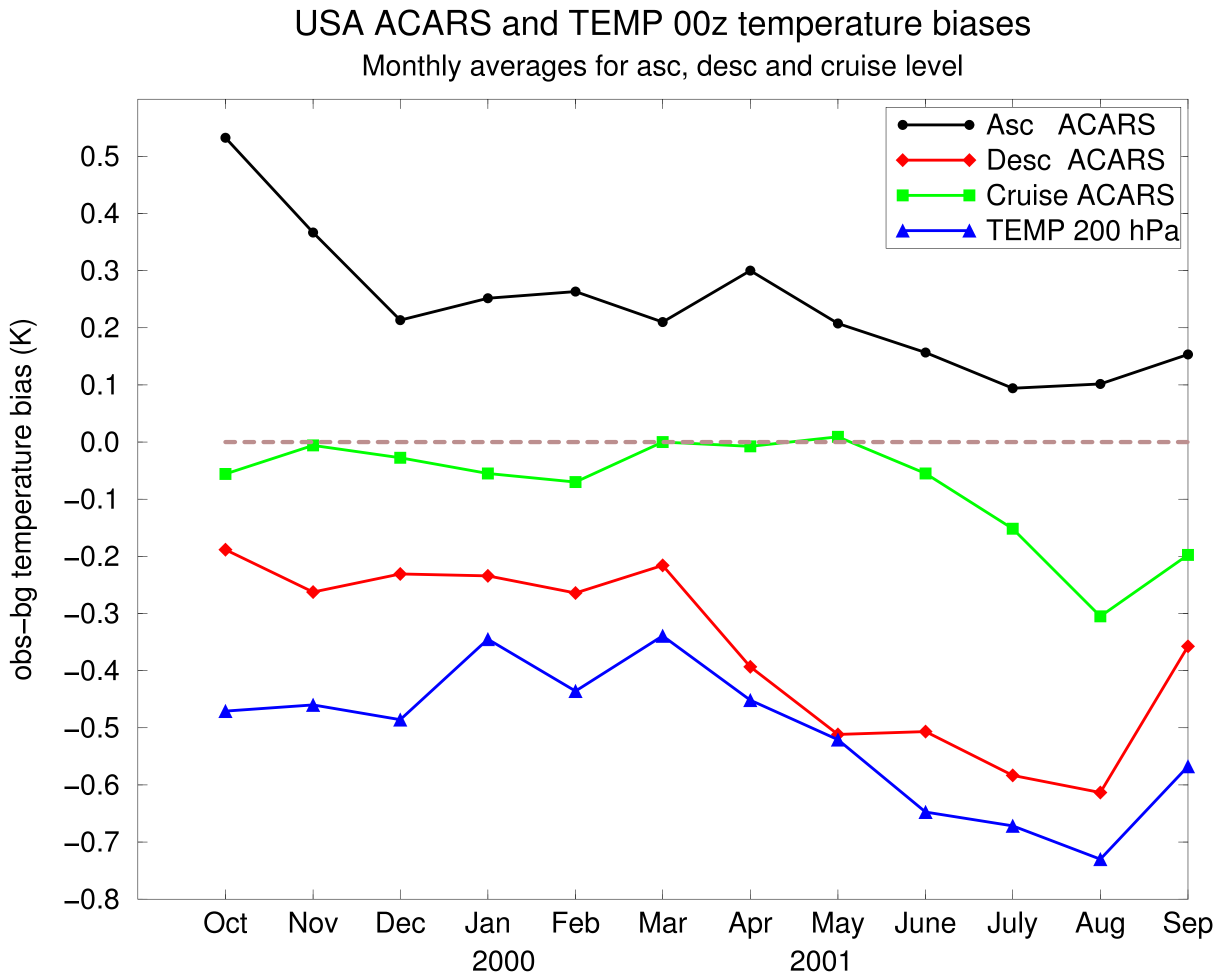



- The only significant source of data in the box is aircraft data (Denver airport).
- The bias for aircraft low level temperature observations was reduced.



ECMWF

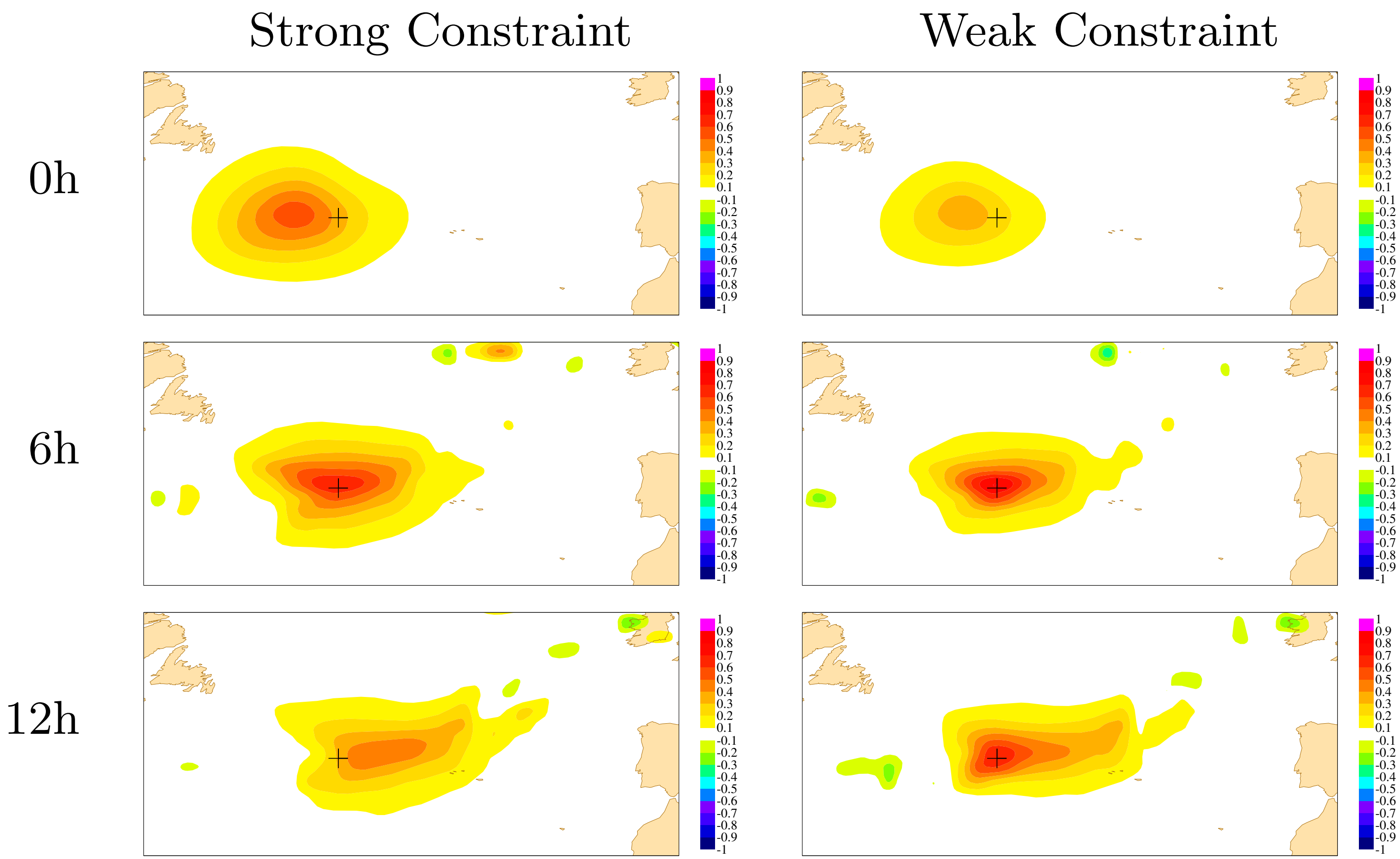
# Aircraft Temperature Bias




Observations are biased.

Figure from Lars Isaksen

# 12 Observations Experiment




- Model error (bias) term captures stationary misfits (model vs. observations).
- The problem to determine the source of the bias remains.



ECMWF

# Conclusions

# Conclusions: Early Results

- With constant model error forcing:
  - Fits the data more uniformly over the assimilation window,
  - Capture some model errors (winter stratosphere),
- With model bias:
  - Improved fit to observations,
  - Bias captures seasonnal variations (to be validated).
- Both capture some observation bias.
- Two tools to define the errors we wish to capture:
  - Model error covariance matrix,
  - Model for model error in $\eta$ and $\beta$ formulations (constant 3D state).

# Conclusions: Weak Constraint 4D-Var

- *Weak constraint 4D-Var* with model bias or forcing model error is essentially an initial value problem with parameter estimation (parameters happen to represent model error or model bias).
- *Weak constraint 4D-Var* with model state control variable is a 4D problem.
  - Takes into account the fact that the model is imperfect without directly estimating model error.
  - *Long window weak constraint 4D-Var* can produce long and consistent 4D pictures of the atmosphere.
  - Model error/bias can be inderectly estimated.
- Applies to operational analysis and reanalysis.

# Conclusions: Future Work

- Long window with 4D model state control variable,
- Compare various formulations of weak constraint 4D-Var ($\eta$, $\beta$, $x$),
- Determine appropriate model error covariance matrix,
- Interactions between model error/bias and observation bias.

- The important questions are:
  - Why are we estimating model error?
  - What type of model error are we estimating (random vs. systematic vs. bias)?